AI-POWERED DISEASE OUTBREAK PREDICTION

LIEW SZE WEN

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS LAPORAN

JUDUL: AI-POWERED DISEASE OUTBREAK PREDICTION

SESI PENGAJIAN: 2022/2023

Saya: LIEW SZE WEN

mengaku membenarkan tesis Projek Sarjana Muda ini disimpan di Perpustakaan Universiti Teknikal Malaysia Melaka dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan unituk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. * Sila tandakan (✓)

_____ SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
_____ TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi / badan di mana penyelidikan dijalankan)
_____√____ TIDAK TERHAD	

Liew
(TANDATANGAN PELAJAR)

Alamat tetap: 21, Jalan Hujung 2,
Sri Lalang, 86000, Kluang, Johor.

Tarikh: 16/6/2023

(TANDATANGAN PENYELIA)

DR. NOOR FAZILLA BINTI ABD YUSOF
Pensyarah Kanan
Jabatan Pengkomputeran Pintar dan Analitik
Fakulti Teknologi Maklumat dan Komunikasi
Universiti Teknikal Malaysia Melaka (UTeM)

Nama Penyelia

Tarikh: 15/9/2023

AI-POWERED DISEASE OUTBREAK PREDICTION

LIEW SZE WEN

This report is submitted in partial fulfillment of the requirements for the Bachelor of Computer Science (Artificial Intelligence) with Honours.

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2022/2023

DECLARATION

I hereby declare that this project report entitled

AI-POWERED DISEASE OUTBREAK PREDICTION

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT : Liew Date : 16/6/2023

(LIEW SZE WEN)

I hereby declare that I have read this project report and found this project report is sufficient in term of the scope and quality for the award of Bachelor of Computer Science (Artificial Intelligence) with Honours.

SUPERVISOR : _____ Date : 15/9/2023

(DR. NOOR FAZILLA ABD YUSOF)

DECLARATION

DEDICATION

I would like to dedicate this project to my parents, who have always supported me financially and emotionally. Throughout this project, I have acquired the ability to tackle complex tasks concurrently effectively and efficiently. Additionally, I express my gratitude to those who have assisted in ensuring the successful completion of this project during its full duration.

ACKNOWLEDGEMENTS

Through this opportunity, my heart extends gratitude to those who helped me in my final year project. Without the guidance and support of Dr Noor Fazilla Abd Yusof, my project would not have been possible. Artificial intelligence expertise proved crucial in reaching the desired outcome.

Furthermore, I would like to thank my parents. Constant mental and motivational support were offered by my parents, which continues to drive me. Keeping me motivated and focused are their encouragement and support.

I would like to express my thanks to my friends. During the completion of the project, friends always provided support and encouragement. The ability to overcome the epidemic is facilitated by mutual aid, technical assistance, and emotional support. Because of their support and continuous encouragement, they play a crucial role in my sphere.

ABSTRACT

The severity of disease outbreaks has increased in recent times. This project aims to use AI techniques, specifically the Long Short-Term Memory (LSTM) algorithm, to predict Monkeypox outbreaks accurately. Traditional reporting methods have limitations, such as being time-consuming, error-prone, and limited scope. Furthermore, language and cultural barriers hinder obtaining precise, detailed outbreak information through manual reporting. An AI-based platform was developed to overcome these challenges, incorporating cultural sensitivity into its computational algorithms. The platform can efficiently analyze vast amounts of data to identify early signs of outbreaks. It provides real-time insights through an analytical dashboard. The LSTM algorithm in this project outperforms other algorithms, with a Root Mean Square Error (RMSE) of less than 1, indicating high precision. The system and dashboard developed through this project are effective tools for health officials to prevent and manage outbreaks proactively. Ultimately, this project aims to enhance disease outbreak prediction and response strategies by utilizing artificial intelligence to mitigate their effects.

ABSTRAK

Wabak penyakit telah menjadi lebih serius sejak tahun-tahun ini. Tujuan projek ini adalah untuk meramalkan wabak penyakit Monkeypox secara tepat dengan menggunakan teknik Kecerdasan Buatan (AI), terutamanya algoritm Long Short-Term Memory (LSTM). Kaedah pelaporan tradisional mempunyai kelemahan seperti penggunaan masa yang panjang, kemungkinan kesilapan dan skop yang terhad. Selain itu, halangan budaya dan bahasa menghalang kakitangan kesihatan daripada mendapatkan maklumat wabak penyakit yang tepat dan terperinci melalui pelaporan manual. Dengan memasukkan kepekaan budaya negara ke dalam algoritm pengiraan, platform berdasarkan AI telah diciptakan untuk mengatasi masalah ini. Platform ini mempunyai keupayaan untuk menganalisis data yang besar untuk mencari gejala awal wabak. Papan pemuka analitik menawarkan pandangan langsung. Dengan kejituan yang tinggi dan ralat Min Kuasa Dua Punca (RMSE) kurang daripada 1, algoritm LSTM dalam projek ini melebihi prestasi algoritm lain. Pihak berkuasa kesihatan boleh menggunakan sistem dan papan pemuka yang dibangunkan dalam projek ini untuk secara proaktif mencegah dan menguruskan wabak penyakit tersebut. Pada akhirnya, AI digunakan dalam projek ini untuk mengurangkan kesan wabak penyakit dan meningkatkan ramalan dan strategi tindak balas.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

LIST OF ABBREVIATIONS

FYP	**-**	**Final Year Project**
LSTM	**-**	**Long Short-Term Memory Network**
RMSE	**-**	**Root Mean Squared Error**
AUC	**-**	**Area Under the Curve**
VS Code	**-**	**Visual Studio Code**

LIST OF ATTACHMENTS

CHAPTER 1: INTRODUCTION

1.1 Introduction

Recently, global diseases have proliferated. COVID-19 has, as we all know, been tormenting us for the past three years. During this time, everyone was housebound. The primary, secondary, and university schools were compelled to close for approximately one year. The employees, however, were encouraged to work from home. This circumstance has made people's existence difficult. In addition, since COVID-19, a few new types of viruses have been discovered.

Early recognition and intervention are essential for limiting and controlling the spread of diseases to prevent viruses from affecting a country's functioning and its citizens' health. Detecting and monitoring diseases can be time-consuming, which increases the likelihood that epidemic outbreaks will spread rapidly.

This initiative utilized AI to predict disease outbreaks to address the issue. Using AI, the system can identify the earliest signs of disease outbreaks. The disease is detectable and treatable more effectively. Early indicators of disease outbreaks can be gleaned from a variety of sources. Examples include electronic health records, social media posts, and news articles. Implementing AI in a system for predicting disease outbreaks can aid public health professionals in automating medical surveillance to obtain more accurate data.

1.2 Problem Statement(s)

- Traditional approaches, such as manual reporting, which are time-consuming, error-prone, and limited in scope, are now used to forecast disease outbreaks.

- Traditional techniques frequently miss early warning signals of disease outbreaks, resulting in slower response times and a greater disease effect.

- Manual reporting techniques might be hampered by language and cultural barriers, making it difficult for health officials to obtain complete and accurate information regarding outbreaks.

1.3 Objective

- To design an AI-powered system capable of analyzing vast amounts of data to detect early signs of disease outbreaks.

- To integrate culturally sensitive algorithms into the AI-powered system to ensure that it can effectively detect disease outbreaks in a variety of groups.

- To develop an analytical dashboard for a decision support system on disease outbreaks and their potential consequences.

1.4 Scope

The development of an AI-powered system for predicting disease outbreaks that integrate data from multiple sources and analyses the data using machine learning algorithms. It provides accurate and timely forecasts of infectious disease outbreaks. The program will concentrate on collecting and integrating data from social media and news articles but will exclude the acquisition of data from environmental sensors and animal populations.

1.5 Project Significance

Because this proposed approach is more precise than current methods, it may be able to detect outbreaks early. Officials in charge of public health and healthcare providers have adequate time to plan for and respond to the outbreak to avoid it becoming more severe. The resource allocation efficiency of public health officials and healthcare providers has dramatically increased. The disease outbreak system powered by artificial intelligence is less expensive and has a smaller impact on the economy than conventional techniques of disease surveillance and outbreak identification.

1.6 Expected Output

This project is expected to yield an accurate, rapid, and secure disease outbreak prediction analytical dashboard powered by AI. The technology will offer public health professionals' early indications of disease outbreaks, allowing them to respond swiftly and efficiently, so minimizing the number of infections and the disease's impact.

1.7 Conclusion

Traditional methods for detecting and accumulating disease information may need to be more efficient and error-prone, delaying responses and aggravating an epidemic. This project seeks to address these issues by creating an AI-powered system for predicting disease outbreaks. The developed AI-powered diseases outbreak prediction system can effectively detect and manage specific disease outbreaks, thereby minimizing their adverse effects on a nation.

It can detect the earliest indicators of an attack by analyzing vast amounts of data. The anticipated outcome of this project is to increase the system's accuracy, which will provide the actual data to public health employees so they can respond swiftly and decide on a series of measures to reduce the number of contaminations and the disease's impact.

CHAPTER 2: LITERATURE REVIEW AND PROJECT METHODOLOGY

2.1 Introduction

Numerous research papers predicting the outbreak of various diseases have been published on the website. To effectively complete the system, it is necessary to employ machine learning to conduct an in-depth analysis of a large quantity of data, including epidemiologic, demographic, and virological data. The primary focus of this review will be on the methods and techniques presented for predicting disease outbreaks. The proposed project methodology will also be explained in this chapter to ensure the project's success.

2.2 Facts and Findings

2.2.1 Domain

The domain of these research papers will be the algorithms and models of applying Artificial Intelligence (AI) and Machine Learning (ML). The developed algorithms and models will be gleaned from relevant research papers to identify the most effective algorithms and models for different diseases, which will then be incorporated into the proposed project.

2.2.2 Existing System

2.2.2.1 Ebola Outbreak Prediction Models

In 2019, there was an epidemic of the virus disease Ebola in the northeastern Democratic Republic of the Congo (DRC). Kelly et al. (2019) uses a non-parametric estimated Hawkes point process model to forecast real-time Ebola outbreaks.

Consequently, the model can estimate the anticipated rate of case accumulation over time and provide real-time outbreak size forecasts. However, the limitations of the model are also discussed. For instance, the inability to parameterize contextual factors, such as contact tracing and clinical care, may affect the trajectory. In addition, disruptions that occur after the predictions may reduce the model's accuracy beyond the uncertainty estimates. In statistical models, parameterizing certain natural biological epidemic factors, such as political unrest or armed conflict, can be challenging.

2.2.2.2 Hand-Foot-Mouth Outbreak Prediction Model

LSTM was discovered by Jia et al. (2019) to be the most suitable model for forecasting hand, foot, and mouth disease (HFMD) outbreaks in China, merging various statistical techniques like Ridge Regression, KNN, Random Forest, and LSTM. The effectiveness of models is evaluated via AUC by looking at Figure 2.1. Particularly in the validation and testing sets, the LSTM model outperforms the other models in predicting HFMD incidence. The limitations of the study are also discussed, particularly the weekly morbidity data absence affecting prediction accuracy. Even though there are difficulties, the research demonstrates that LSTM could potentially be applied to the analysis of infectious disease morbidity and the forecasting of future events.